

APPROVED: 24 January 2020

doi:10.2903/sp.efsa.2020.EN-1787

## Validation and inference of high-resolution information (downscaling) of ENETwild abundance model for wild boar

ENETWILD-consortium<sup>1</sup>, Pelayo Acevedo<sup>2</sup>, Simon Croft<sup>3</sup>, Graham C Smith<sup>3</sup>, Jose Antonio Blanco-Aguilar<sup>2</sup>, Javier Fernandez-Lopez<sup>2</sup>, Massimo Scandura<sup>4</sup>, Marco Apollonio<sup>4</sup>, Ezio Ferroglio<sup>5</sup>, Oliver Keuling<sup>6</sup>, Marie Sange<sup>6</sup>, Stefania Zanet<sup>5</sup>, Francesca Brivio<sup>4</sup>, Tomasz Podgórski<sup>7,8</sup>, Karolina Petrovič<sup>7</sup>, Ramon Soriguer<sup>2</sup>, Joaquín Vicente<sup>2</sup>

<sup>1</sup>ENETWILD Consortium: [www.enetwild.com](http://www.enetwild.com), <sup>2</sup>National Institute on Wildlife Research (IREC), University of Castilla-La Mancha and Consejo Superior de Investigaciones Científicas, Ciudad Real, Spain. <sup>3</sup>National Wildlife Management Centre, Animal and Plant Health Agency, Sand Hutton, York, UK. <sup>4</sup>Department of Veterinary Medicine, University of Sassari, Sassari, Italy. <sup>5</sup>University of Torino, Torino, Italy. <sup>6</sup>Institute for Terrestrial and Aquatic Wildlife Research, University of Veterinary Medicine Hannover, Hannover, Germany. <sup>7</sup>Mammal Research Institute, Polish Academy of Sciences, Białowieża, Poland. <sup>8</sup>Czech University of Life Sciences, Prague, Czech Republic.

### Abstract

The ENETWILD consortium provided in August 2019 a map at 10x10 km resolution for wild boar abundance based on hunting data. The availability of prediction maps at a spatial resolution comparable with the one of the home range of wild boar can be useful for further evaluation of risk of spread of African swine fever (ASF). Therefore, predictions of abundance on the basis of the wild boar home range are required. The downscaling procedure needs information on what resolution level is being used for predictions (hunting grounds, municipalities and NUTS3). This report presents the validation of previously produced hunting yield maps (10x10 km resolution) and new model projections downscaled at 2x2 km resolution. A new dataset based on hunting bag numbers was used as external data for validation. These data were arranged at two levels: at country level for the European scenario and at NUTS3 level for a scenario in Spain, where the data availability is higher than the rest of Europe in terms of quantity and quality. Very similar geographical patterns of wild boar abundance were obtained when the models were transferred to 2x2 km grid. The downscaled model predictions were aggregated at country and NUTS3 levels and compared against the external dataset. Our study confirmed that both 10x10 km and 2x2 km resolutions were able to detect spatial variation in wild boar hunting bags (high model performance) and to predict the numbers of wild boar hunted with relative precision (moderate model accuracy). Nevertheless, an overestimation of absolute number of hunted wild boar was observed using both resolutions. Reasons for this overestimation are discussed in this report. The linearity between predictions of hunting yield and external dataset was maintained, indicating that hunting yield predictions can be considered as a good proxy of wild boar abundance. Therefore, updated wild boar hunting yield data, collected at the finest spatial resolution as possible, is needed to correctly recalibrate our model at regional level, in particular in eastern European countries.

© European Food Safety Authority, 2020

**Key words:** distribution, game management, hunting bags, population abundance, population monitoring, risk assessment, spatial modelling, *Sus scrofa*, wild boar, downscaling

**Question number:** EFSA-Q-2019-00249

**Correspondence:** [alpha@efsa.europa.eu](mailto:alpha@efsa.europa.eu)

**Disclaimer:** The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**Acknowledgements:** We acknowledge EFSA ALPHA and DATA units, and ENETWILD partners for reviewing this manuscript. We are grateful to ENETWILD collaborators (listed on <http://www.enetwild.com/collaborators>) and data providers: C. Fonseca (Portugal), A. Amici (Italy), G. De Castro (Italy), R. Fontana (Italy), M. Bonacoscia (Italy), V. Daković (Montenegro), G. Markov (Bulgaria), P. Genov (Bulgaria), A. Licoppe (Belgium), J. Casaer (Belgium), A. Licoppe (Belgium), C. Malengreaux (Belgium), M. de Tillesse (Belgium), S. Cellina (Luxembourg), W. Schweiz (Switzerland), S. Meier (Switzerland), R. Solène (France), N. Sprem (Croatia), K. Kavčić (Croatia), B. Pokorny (Slovenia), D. Gačić (Serbia), O. Belova (Lithuania), I. Zihmanis (Latvia), S. Cellina (Luxemburg), R. Veeroja (Estonia), A. Chaikouski (Belarus), V. Shakun (Belarus), R. Phlal (Czech Rep.), M. Jezek (Czech Rep.), J. Bučko (Slovakia) and institutions feeding GBIF (listed at <https://doi.org/10.15468/dl.v2ydsu>) for sharing data on wild boar and other species.

**Suggested citation:** ENETWILD consortium. Validation and inference of high-resolution information (downscaling) of ENETwild abundance model for wild boar. EFSA supporting publication 2020:EN-1787. 23 pp. doi:10.2903/sp.efsa.2020.EN-1787

**ISSN:** © European Food Safety Authority, 2020

## Summary

Downscaling is the procedure to infer high-resolution information from coarse resolution sources. It is commonly applied in models for wild animal distributions using fine-grain environmental data to predict species' distributions and abundances at high resolution and offers great potential benefit for ecological and epidemiological studies. The ENETWILD consortium recently presented a model of wild boar relative abundance based on hunting yields for MSs and neighbouring countries. The model outputs were presented at 10x10 km as a standard resolution used in many national mammal atlas. As the data used in model calibration was provided at three different resolution levels (hunting grounds, municipalities and NUTS3) the downscaling procedure requires information of the resolution level that is used for predictions. Therefore, 10x10 km cells were assimilated to hunting ground level in a first downscaling process. The availability of prediction maps at a spatial resolution comparable with the size of home range of wild boar (i.e. the typical area over which one individual or a group of wild boar move which overlaps with those of neighbouring animals or groups) are useful to further study the risk of spread of African swine fever (ASF) in wild boar populations. This is because contacts among individuals or groups are likely to be influenced by the number of individuals with neighbouring (adjacent) home ranges. In this report we present a further step in modelling the relative abundance of wild boar based on hunting yield data by: i) assessing 10x10 km grid predictions assuming cell area similar to both hunting ground level (from previous report) and similar to municipality level (new predictions) and ii) exploring the possibility to downscale the model predictions into a new 2x2 km grid (approaching the size of an average wild boar home range).

New datasets based on hunting bag numbers were provided independently by governmental administrations and were used as external datasets for model validation. The data were arranged at two levels: country level for European scale and NUTS3 for Spain. Then, model predictions at 10x10 km were aggregated at the same level as external dataset: at country level for the European scenario and at NUTS3 level for Spain. Subsequently, these aggregated values were compared to observed data in order to assess the predictions performance and accuracy.

The model predictions downscaled at 10x10 km showed a linear relation to external hunting bag data. High model performance was obtained when assuming both, cells as hunting ground (previous report) as well as cells as municipalities (current report). In general, our downscaled model predictions presented an overestimation compared to values from external databases. These overpredicted hunting yields could be closer to the actual wild boar abundance, although more wild boar densities data are still required to test these relationships. While data at municipality level reports a total amount of wild boar hunted in an administrative area which may intrinsically include zones where hunting is not allowed, in downscaling processes cells occupy the totality of the territory assuming that hunting is allowed in the whole administrative area. Therefore, we consider that predictions assuming cells as municipalities would be closer to the actual wild boar abundance rather than a prediction of hunting yields.

Predictor variables at original resolution encompassed the range observed at 2x2 km grid (except for North Europe and other few locations at East Europe), and the relation among them was maintained when downscaling. Changes between bioregions in the predicted values were more abrupt at this resolution than for 10x10, probably due to the interaction AREA\*BIOREGION. Model predictions at 2x2 km assuming cell as hunting ground performed well following the evaluations on hunting bag external data, which corroborate that the model maintain their predictive power at different scales. This will be especially important in the modelling of abundance for other species which may require even finer resolutions.

Model evaluation also showed differences in model performance depending on the country. In particular, hunting bags were largely overpredicted in a significant group of countries around Balkans (Serbia, Bulgaria, Romania and Ukraine). This could be due to under-reporting of hunting activities in those countries; so more and better data is needed to improve and validate the

models for this area. Nevertheless, we cannot exclude the possibility that, overall, wild boar populations have potential for further increasing their abundance (or to a lesser extent compared with other parts of Europe) in this region with such a highly favourable habitat.

The use of this model species to downscale the resolution of distribution map predictions evidenced methodological issues, which will be very valuable for future modelling of wildlife species at a continental level. The importance in taking into account the heterogeneity in model input data or the possibility to use different model parameterization for each bioregion are some of the approaches to explore in the next activities of ENETWILD.

## Table of contents

Abstract .....	1
Summary .....	3
1. Introduction .....	6
1.1. Background and Terms of Reference as provided by the requestor .....	6
1.2. Scope of the report .....	6
2. Data .....	8
2.1. Study area .....	8
2.2. Environmental variables and other predictors .....	9
3. Methodology .....	11
3.1. Evaluation of the hunting yield model predictions at 10x10 km grid .....	11
3.2. Downscaling approach at 2x2 km grid and model evaluation .....	11
4. Results and discussion.....	12
4.1. External evaluation of predictions at 10x10 km from hunting yield model .....	12
4.2. Downscaling at 2x2 km and evaluation.....	15
4.3. Conclusions and further steps.....	18
References.....	20
Glossary .....	21

## 1. Introduction

### 1.1. Background and Terms of Reference as provided by the requestor

This contract was awarded by EFSA to Universidad de Castilla-La Mancha, contract title: Wildlife: collecting and sharing data on wildlife populations, transmitting animal disease agents, contract number: OC/EFSA/ALPHA/2016/01 – 01.

The terms of reference for the present report were to model wild boar distribution based on hunting yields to apply further efforts in validating previous predictions at 10x10 km resolution and make first steps to transfer hunting yield to a higher 2x2 km resolution.

### 1.2. Scope of the report

The ENETWILD consortium ([www.enetwild.com](http://www.enetwild.com)) implemented an EFSA funded project whose current main objective is to collect information regarding the geographical distribution and abundance of wild boar throughout Europe in order to subsequently create geospatial tools to be used in further risk assessment of diseases such as ASF. ENETWILD consortium has provided several spatial distribution model updates with new data from gap areas and continuously verifying and validating the methodological approach. The table 1 summarizes the specific objectives, model settings and predictors, and the main conclusions of the previous reports that guided further steps in the modelling process.

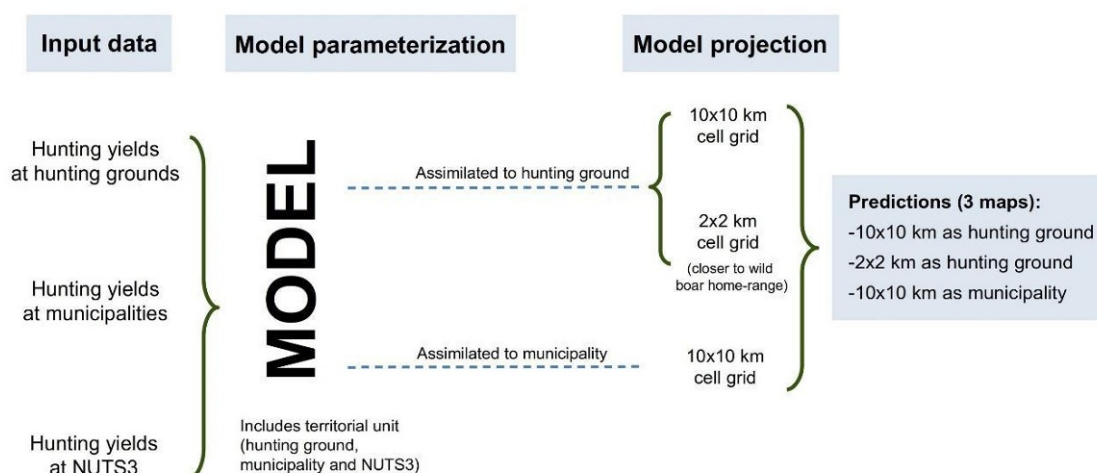
**Table 1:** Summary of main aims and outcomes of the last reports provided by ENETWILD consortium regarding spatial distribution models of wild boar in Europe. We indicate the specific objectives, model settings and predictors, and the main conclusions that guided further steps in the modelling process for abundance (hunting) data.

Report	Objetives	Model settings and predictors	Conclusions guiding future methodological steps
ENETWILD modelling of wild boar distribution and abundance: initial model output based on hunting data and update of occurrence-based models (April 2019)	<ul style="list-style-type: none"> <li>- Prediction of wild boar relative abundances</li> <li>- Downscaling to 10x10 km grid</li> </ul>	<ul style="list-style-type: none"> <li>- Bioclimatic and land cover</li> <li>- Altitude, latitude, longitude</li> <li>- Habitat suitability (Alexander et al. 2016)</li> <li>- Bioregion (Pittiglio et al. 2018)</li> </ul>	<ul style="list-style-type: none"> <li>- To increase model sample size and fill gap areas</li> <li>- To test new predictors to improve predictions</li> <li>- To explore the relation between resolution of input data and fine grain predictions</li> </ul>

<p>ENETWILD modelling of wild boar distribution and abundance: update of occurrence and hunting data-based models (August 2019)</p>	<ul style="list-style-type: none"> <li>- Incorporation of new predictor variables</li> <li>- Exploration of relations among bioregion, resolution of input data and area</li> <li>- New downscaling to 10x10 km grid</li> </ul>	<ul style="list-style-type: none"> <li>- Altitude, sun radiation, human influence, snow cover, vegetation growing period, area</li> <li>- Bioregion (ENETWILD consortium)</li> <li>- NUT (resolution level of input data)</li> <li>- Interaction between area and NUT</li> <li>- Interaction between area and bioregion</li> <li>- Assuming cells as hunting grounds in 10x10 km grid downscaling</li> </ul>	<ul style="list-style-type: none"> <li>- To increase model sample size at high resolution level (hunting grounds)</li> <li>- To explore alternative modeling approaches such as predictions at 2x2 km grid</li> <li>- To parametrize independent models for each bioregion if more data is available</li> <li>- To obtain independent data for model evaluation</li> </ul>
<p>Validation and downscaling of ENETWILD abundance model for wild boar (present report)</p>	<ul style="list-style-type: none"> <li>- Validation of previously produced hunting yield maps and new ones</li> <li>- Downscaling to 10x10 km grid</li> <li>- Downscaling to 2x2 km grid</li> </ul>	<ul style="list-style-type: none"> <li>- Model from ENETWILD report August 2019</li> <li>- Assuming cells as municipality in 10x10 km grid downscaling</li> <li>- Assuming cells as hunting grounds in 2x2 km grid downscaling</li> </ul>	<ul style="list-style-type: none"> <li>- To update wild boar hunting yield data for some specific regions</li> <li>- To increase hunting yield data resolution</li> <li>- To explore model independent parametrization for each bioregion</li> </ul>

ENETWILD recently presented a model of wild boar hunting yield predictions from regional and local hunting records (ENETWILD consortium et al. 2019). The availability of prediction maps at a spatial resolution comparable with the size of home range of wild boar (i.e. the typical area over which one individual or a group of wild boar move which overlaps with those of neighbouring animals or groups, Woodroffe 1999; Barasona et al. 2014; Pepin et al. 2016) can be useful for further evaluation of risk for ASF spread. This is because contacts among individuals or groups are likely to be influenced by the number of individuals with neighbouring (adjacent) home ranges. There is great potential value and opportunity in the downscaling of species distributions, that is, the combination of coarse-grain species data (distribution, abundance) with fine-grain environmental data to predict species' distributions at a fine grain (Keil et al. 2012). However, the quality of results provided by downscaling methods must be always assessed.

Once obtained the hunting yield model and downscaled at 10x10 km grid, the goal of this report is to evaluate the performance of the previous 10x10 km grid predictions with external data (provided independently by administrations and not used in model calibration) and to carry out a new downscaling of the model to produce predictions both to 10x10 km, assuming a different relationship between the number of hunted boars and area (see above), and into a finer resolution (2x2 km cells)(Figure 1). The new resolution was chosen according to the wild boar home range, that is closer to a 4 km<sup>2</sup> scale than to a 100 km<sup>2</sup> one (e.g. Massei et al. 1997; Keuling et al. 2008). Finally, we identify the weak points of the models and establish the roadmap for future modelling.



**Figure 1:** Flowchart of the hunting yield modelling (see ENETWILD consortium et al. 2019 for model parametrization details) and projection process.

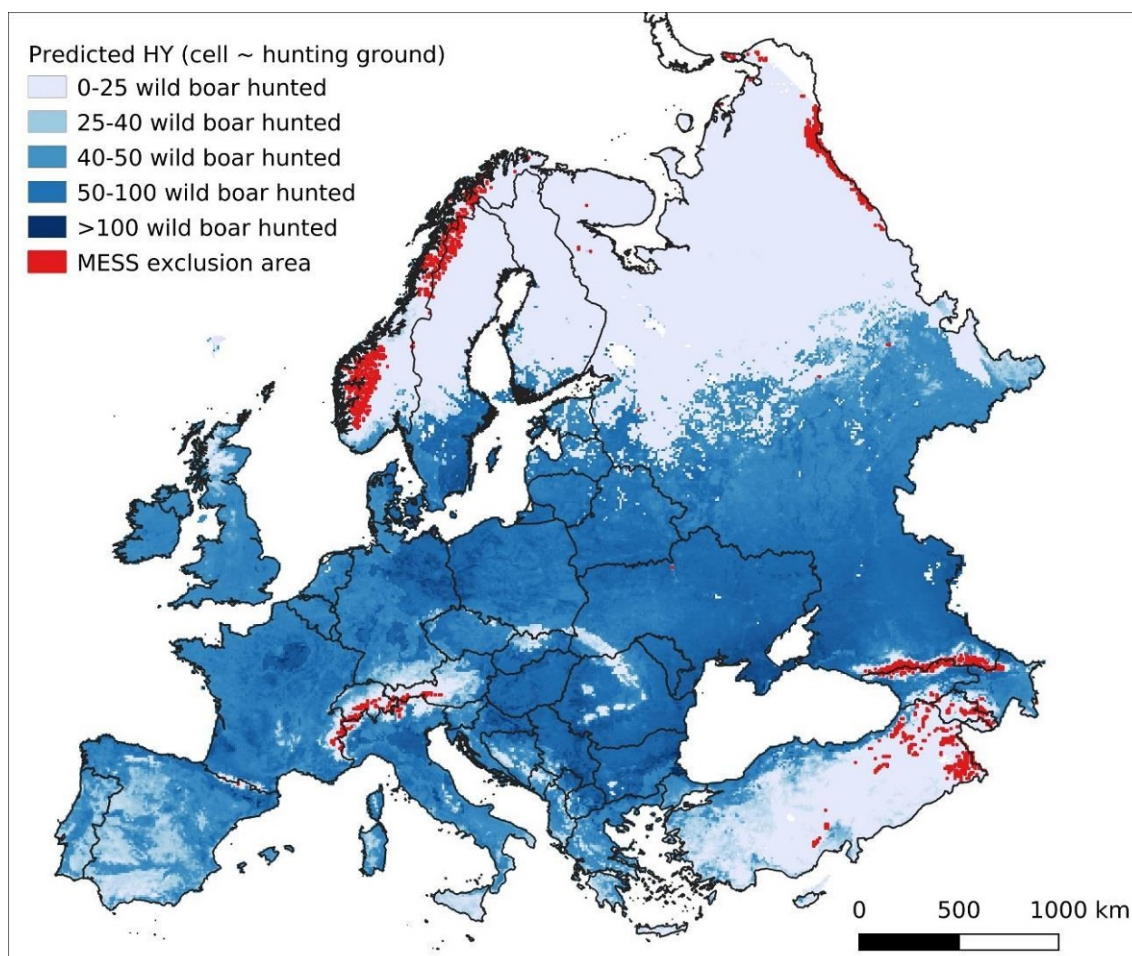
Model was parameterized in the previous report using different types of input data. The model, therefore, includes a factor defining the input data (hunting ground, municipalities and NUTS3) and its interaction with surface among the predictors, meaning that a different relationship between surface and hunting yields is observed. Therefore, when this model has to be projected into new territorial units (cells), it should be specified the type of the input data to which the cells are assimilated, in surface terms. In the previous report, model was downscaled to 10x10 km grid (cell ~ hunting ground). In the current report, model downscaling to 2x2 km grid (cell ~ hunting ground) and to 10x10 km grid (cell ~ municipality) were addressed.

## 2. Data

### 2.1. Study area

The study area is the same as the one used in the previous report (ENETWILD consortium et al. 2019) and spans 11,019,700 km<sup>2</sup>. It comprises all countries in mainland Europe approximately delimited by the Ural Mountains at the Eastern boundary which is also likely to act as a geographical barrier limiting the immigration of wild boar from Asia (IUCN map on wild boar distribution at <https://www.iucnredlist.org/species/41775/10559847>), and includes Mediterranean islands and the UK and Ireland (Figure 2). For this report the study area was divided into 2x2 km grid cells, resulting in 2,787,877 cells that were used in model downscaling processes (see below).





For the predictions 10x10 km squares were assimilated to hunting ground level (see text for details; see also Figure 9 in the previous report, ENETWILD consortium et al. 2019). Red squares represent localities beyond the environmental domain of the model according to Multivariate Environmental Similarity Surface analysis (Elith et al. 2010).

**Figure 2:** Extent of the study area and predicted hunting yield (HY; number of wild boar hunted) from the simplified final model (Equation 1; see also ENETWILD consortium et al. 2019) at 10x10 km EEA grid.

## 2.2. Environmental variables and other predictors

Predictor variables used in the last report to model wild boar hunting bags were resampled to 2x2 km resolution from the original raw sources (Table 2; see also ENETWILD consortium et al. 2019). This new grid was obtained by splitting the original 10x10 km provided by the European Environment Agency (<https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2>). Briefly, bioclimatic variables and sun radiation were obtained from the Worldclim 2 project database. Land use data came from ESA/CCI-LC project, version v2.0.7 (2015) (<https://www.esa-landcover-cci.org/?q=node/158>). Mean altitude was extracted from the USGS Space Shuttle Radar Topography Mission (SRTM) GL30 (<https://lta.cr.usgs.gov/SRTM1Arc>) and snow cover was obtained from MODIS/Terra Snow Cover project (Monthly L3 Global 0.05Deg CMG, Version 6; <https://nsidc.org/data/MOD10CM>). Human footprint index (an index related to human population distribution, urban areas, roads, etc.) was provided by The Last of the Wild Project version 2 (<http://sedac.ciesin.columbia.edu/data/collection/wildareas-v2>). Vegetation growing period is based on a water balance model (<http://www.appsolutelydigital.com/DataPrimer/part154.html>). In addition, each 2x2 km cell was assigned to a bioregion according to the regionalization

previously reported (ENETWILD consortium et al. 2019). Raster predictor layers and 2x2 km grid polygons were managed using QGIS 3.4 and “rgdal” R package (Bivand et al. 2019).

**Table 2:** Variables used to model the spatial distribution of wild boar abundance and distribution at European scale. All of them were managed and resampled at 2x2 km resolution.

<i>Code</i>	<i>Variable description</i>	<i>Code</i>	<i>Variable description</i>
BIO1	Annual mean temperature	lc_10	Cropland, rainfed
BIO2	Mean diurnal range (mean of monthly (max temp - min temp))	lc_11	Herbaceous cover
BIO3	Isothermality (BIO2/BIO7) (x 100)	lc_12	Tree or shrub cover
BIO4	Temperature seasonality (SD x 100)	lc_20	Cropland, irrigated or post-flooding
BIO5	Max temperature of warmest month	lc_30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)
BIO6	Min temperature of coldest month	lc_40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)
BIO7	Temperature annual range (BIO5-BIO6)	lc_60	Tree cover, broadleaved, deciduous, closed to open (>15%)
BIO8	Mean temperature of the Wettest Quarter	lc_61	Tree cover, broadleaved, deciduous, closed (>40%)
BIO9	Mean temperature of the Driest Quarter	lc_70	Tree cover, needleleaved, evergreen, closed to open (>15%)
BIO10	Mean temperature of warmest quarter	lc_71	Tree cover, needleleaved, evergreen, closed (>40%)
BIO11	Mean temperature of coldest quarter	lc_80	Tree cover, needleleaved, deciduous, closed to open (>15%)
BIO12	Annual precipitation	lc_90	Tree cover, mixed leaf type (broadleaved and needleleaved)
BIO13	Precipitation of wettest month	lc_100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)
BIO14	Precipitation of driest month	lc_110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)
BIO15	Precipitation seasonality (coefficient of variation)	lc_120	Shrubland
BIO16	Precipitation of wettest quarter	lc_122	Deciduous shrubland
BIO17	Precipitation of driest quarter	lc_130	Grassland
BIO18	Precipitation of Warmest Quarter	lc_140	Lichens and mosses
BIO19	Precipitation of Coldest Quarter	lc_150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)
GROW	Length of vegetation growing period	lc_152	Sparse shrub (<15%)
SUNRAD	Sun radiation	lc_153	Sparse herbaceous cover (<15%)
SNOW	Snow cover	lc_160	Tree cover, flooded, fresh or brakish water
HFP	Human Footprint Index	lc_180	Shrub or herbaceous cover, flooded, fresh/saline/brakish water
NUT	Administrative level (categorical: hunting ground, municipality and NUTS 3)	lc_190	Urban areas
ASF	Years since the first ASF report	lc_200	Bare areas
BIOREG	Environmental bioregions (categorical: North, South, East and West)	lc_201	Consolidated bare areas
ALT	Mean altitude	lc_202	Unconsolidated bare areas
AREA	Area of sampling unit	lc_210	Water bodies
		lc_220	Permanent snow and ice

### 3. Methodology

#### 3.1. Evaluation of the hunting yield model predictions at 10x10 km grid

In the last report (ENETWILD consortium et al. 2019), a generalised linear model was used to predict wild boar hunting yields. Original wild boar hunting data (observed number of animal hunted yearly) were obtained from a range of administrative units, from hunting grounds to NUTS3 level or equivalent (input data in Figure 1). The model was parameterized using the maximum number of wild boar hunted annually within 2014-2018 hunting seasons as a response variable with a negative binomial distribution (to control the overdispersion in the data) and a logarithmic link function. As the data used in model parameterization was provided at three levels (hunting grounds, municipalities and NUTS3), a predictor factor was included in the model to indicate the resolution level of each sample unit (called NUT). The simplified version of the model was obtained using the information of the stepwise procedure, based on the reductions in AIC values of the predictors (see Equation 1). Predictions were internally and successfully validated against a subset of the original data (20% of the total records, a common percentage used in training/validation partitions following the Pareto Principle, Box and Meyer 2012).

$$HY = \exp(3.302 + (9.801e-09*AREA) + (2.044*municipality) + (3.742*NUTS3) + (5.974e-01*BIOREG2) + (1.425*BIOREG3) + (6.278e-01*BIOREG4) + (-1.685e-02*BIO15) + (-4.510e-04*ALT) + (-6.362e-03*HFP) + (-3.262e-02*SNOW) + (1.248e-02*lc_70) + (1.075e-02*lc_60) + (9.077e-03*lc_120) + (5.257e-03*lc_10) + (3.519e-03*lc_100) + (-9.004e-09*AREA:municipality) + (-9.894e-09*AREA:NUTS3) + (4.486e-10*AREA:BIOREG2) + (1.483e-10*AREA:BIOREG3) + (3.447e-10*AREA:BIOREG4))$$

[Equation 1]

Once validated, the model was downscaled to a 10x10 km grid. This procedure requires information about what resolution level (hunting grounds, municipalities or NUTS3) is being used for predictions as it was a predictor factor included in the final model as well its interaction with surface (variable AREA). Thus, 10x10 km cells were assumed as hunting grounds to define the relationship between the variables AREA and hunting yield (see Figure 8 in ENETWILD consortium et al. 2019). In this report, the model predictions are evaluated against external data: number of wild boar hunted by hunting season provided independently by governmental administrations. These data were arranged at two different administrative levels, namely countries at European level and NUTS3 for Spain. To assess the performance of the wild boar model downscaled to a 10x10 km grid, we used different predictions: i) assuming cells as hunting grounds (previous report) and ii) assuming cells as municipalities. In most cases this last approach could be more realistic, since 10x10 km cells are more similar to municipalities than to hunting ground in terms of size. Predictions from both projections were aggregated according to external datasets (countries at European level and NUTS3 for Spain). Then, the relationships between the predicted number of wild boar hunted and the numbers reported in the external datasets was assessed by using Pearson's correlations.

#### 3.2. Downscaling approach at 2x2 km grid and model evaluation

In order to obtain finer resolution predictions of the wild boar hunting yield model, the simplified model described in Equation 1 (more details in ENETWILD consortium et al. 2019), was projected to a 2x2 km grid. In order to assess the transferability of the model to the 2x2 km grid, we followed the analytical procedure suggested by Werkowska et al. (2017). Correlation matrices among predictors at different resolution were performed, to ensure that the correlation pattern between predictors is maintained from the original territorial units to the 2x2 km grid. In addition, Multivariate Environmental Similarity Surfaces (MESS) analyses (Elith et al. 2010) were conducted to tests if the range of the predictor variables at the training data encloses the range of the

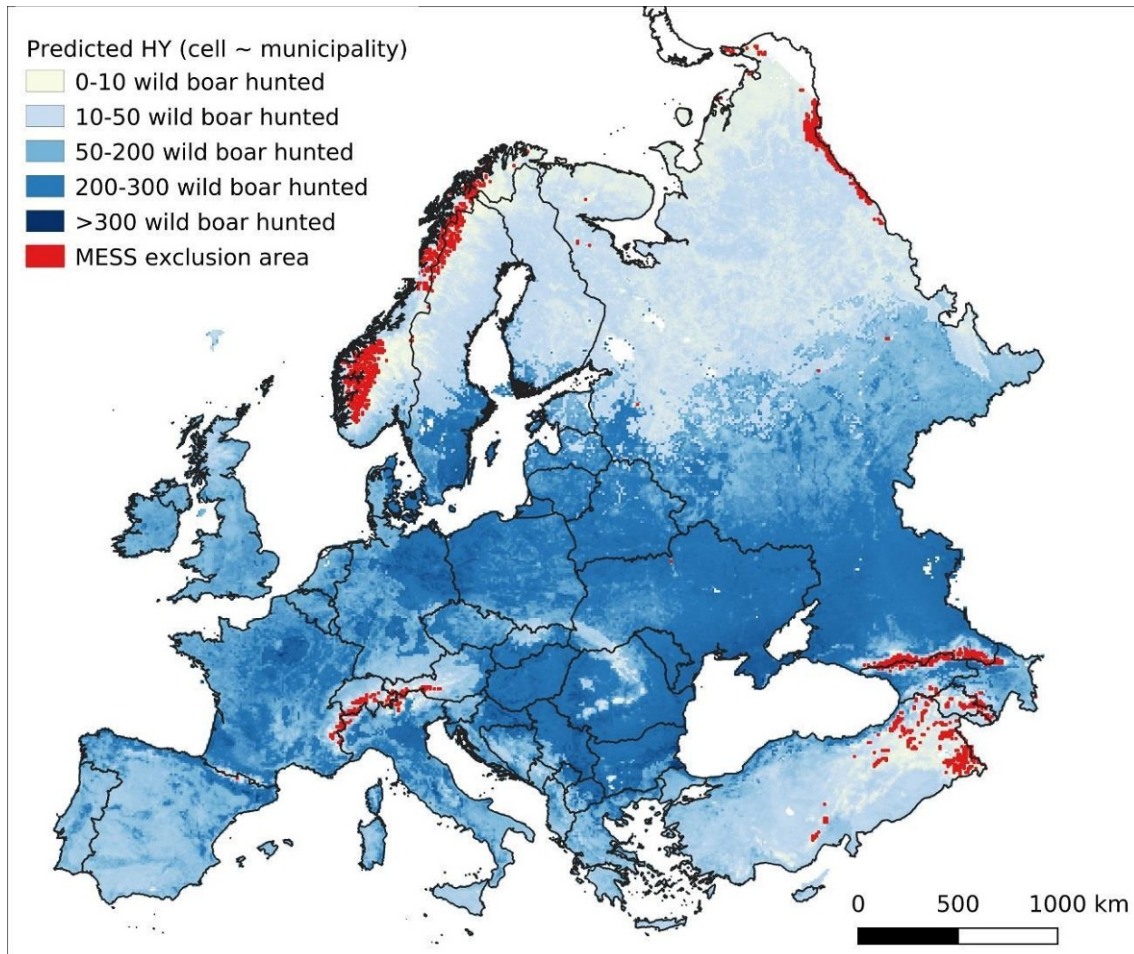
variables at the projection scale (2x2 km grid in our case). The variable AREA as predictor was set at 4,000,000 m<sup>2</sup> for all cells in the 2x2 km grid, and variable NUT was established as hunting ground level since the mean size of this type of territorial unit (30 ± (SD) 44 km<sup>2</sup>) was the most similar to the cells area.

Downscaled 2x2 km predictions were assessed against external data in the same way as the 10x10 km grid predictions. Cell predictions aggregated as a sum for each territorial unit were compared to their relatives in the external validation datasets by using Pearson's correlations.

## 4. Results and discussion

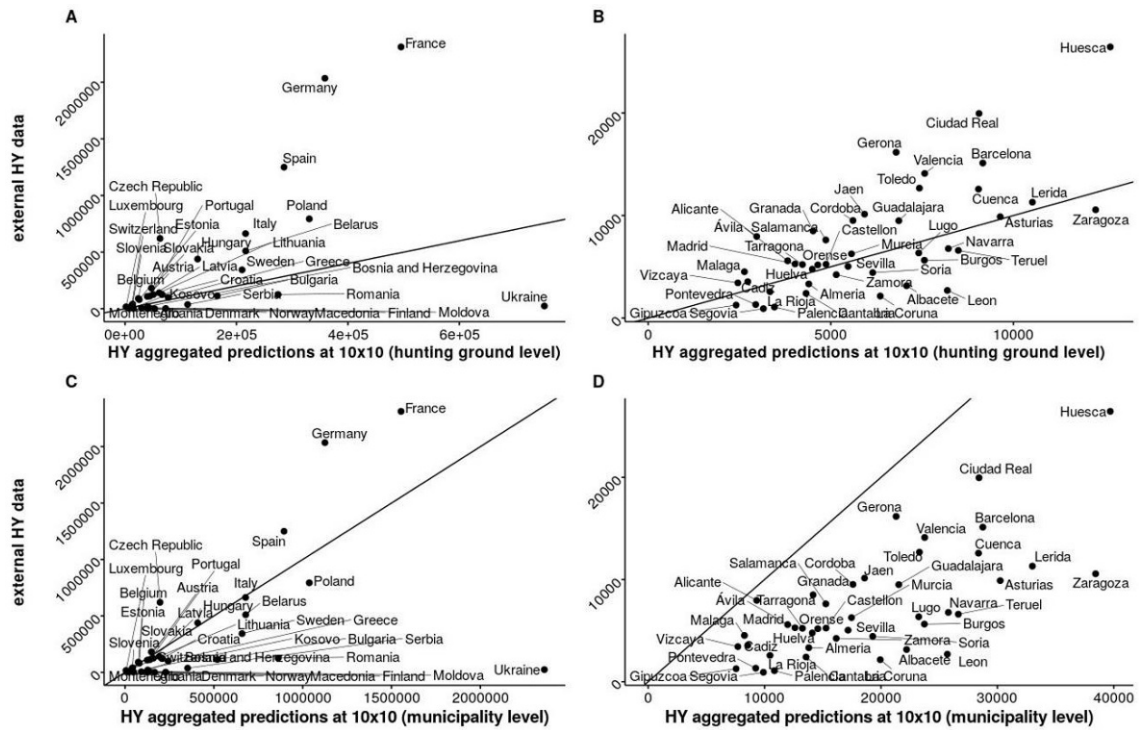
### 4.1. External evaluation of predictions at 10x10 km from hunting yield model

In general, the same spatial pattern of hunting yield predictions were obtained at 10x10 km when assuming cell size to reflect hunting ground or municipality (Figures 2 and 3). Model predictions were positively correlated to external hunting yields numbers for countries ( $R=0.57$ ;  $p < 0.01$ ) and NUTS3 ( $R=0.71$ ;  $p < 0.01$ ). This relationship is the same in both approaches, when assuming cell size to reflect hunting ground and municipality (Figure 4). However, absolute numbers of wild boar hunted varied between approaches, showing higher hunting yield predictions when cells were assimilated as municipalities (Figure 4). This approach fits best the data at country level, but slightly overestimate hunting yields at NUTS3 for Spain. However, overpredictions are more reliable for downscaling projections. In our model, hunting yield input data at municipality level reports a total amount of wild boar hunted in an administrative area which presumably includes zones where hunting is not allowed or temporally banned. In downscaling processes, cells occupy the totality of the territory, assuming that hunting is available in the whole administrative area (countries or NUTS3). For this reason, although both model outputs can be used as a proxy of wild boar relative abundance, we consider predictions assuming cells as municipalities closer to the actual wild boar abundance, therefore this approach should be followed.



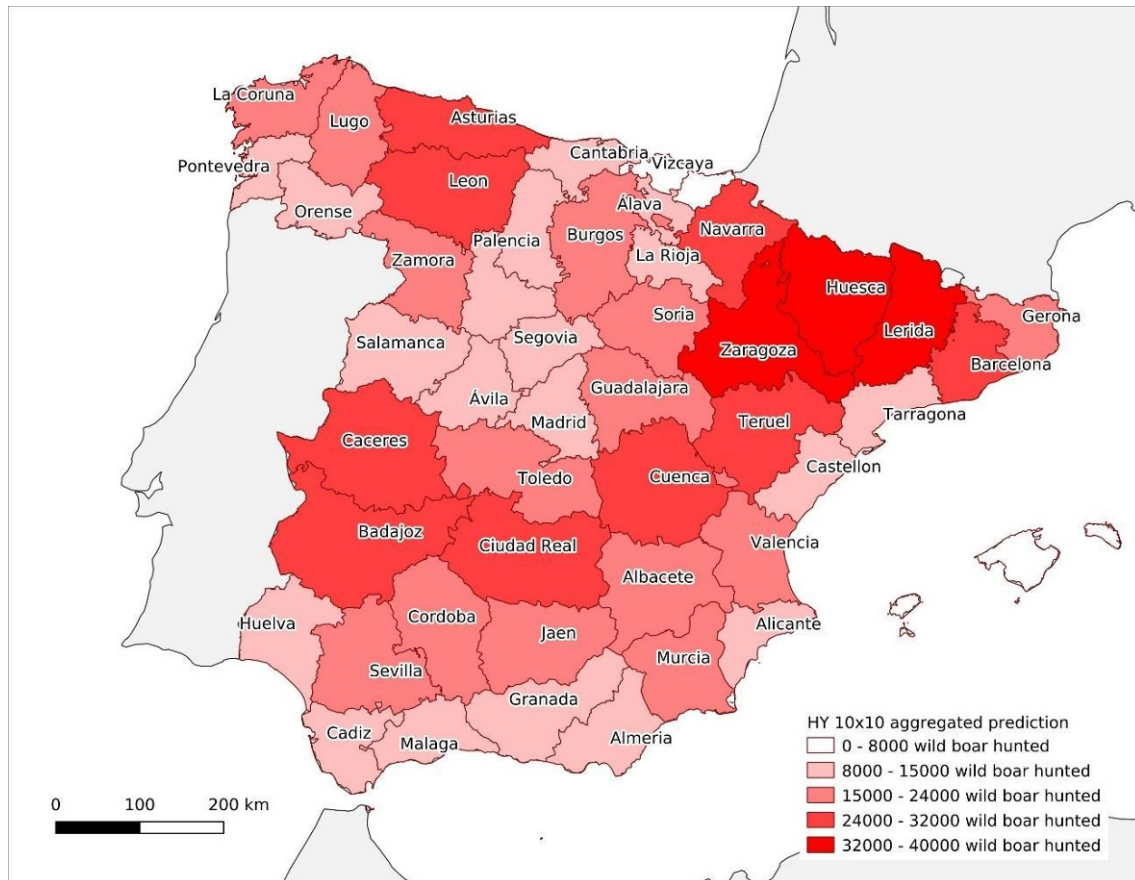
**Figure 3:** Predicted hunting yield (HY; number of wild boar hunted) from the simplified final model (Equation 1; see also ENETWILD consortium et al. 2019) at 10x10 km EEA grid (squares were assimilated to municipality, notice that the legend numbers differ from those in Figure 2, see text for details). Red squares represent squares beyond the environmental domain of the model according to Multivariate Environmental Similarity Surface analysis (see Elith et al. 2010).

For datasets at country level (cell ~ municipality), three different groups could be distinguished: i) a main group for which our model accurately fits well validation hunting dataset (group A; Figure 4-C, bottom left), ii) a second group of countries for which external data showed higher values than obtained from predictions at 10x10 km grid (group B; Czech Republic, Spain, Germany and France; Figure 4-C), and iii) a group of countries where wild boar hunting is strongly overpredicted (group C; Serbia, Bulgaria, Romania and Ukraine; Figure 4-C). Group B includes countries with a long hunting tradition, where, among other management related factors, supplementary feeding at hunting grounds could explain higher values in comparison to our predictions. On the other hand, Group C mainly encompasses areas where wild boar is reaching (i.e. wild boar expansion areas), and other countries around the Balkans zone. However, this spatial aggregation of countries around the Balkans could indicate a lack of good quality data for this area (hunting is underreported), so a higher effort should be made to obtain more precise data of hunting animal for this region in near future. We cannot exclude the possibility that, overall, wild boar populations still have not displayed all its population dynamics potential (or to a lesser extent compared with other parts of Europe) in such a highly favourable habitat region.



**Figure 4:** Relationships between aggregated model predictions at 10x10 km grid cells at different administrative levels and external validation datasets. A and B assumes cells size to hunting grounds, since C and D assumes cells as municipalities. Left plots correspond to validations at European countries level, while right plots represent validations at Spanish NUTS3 level. Lines show the identity ( $y=x$ ).

At NUTS3 level in Spain (as an example), when assuming cell size to reflect municipality (Figure 4-D and Figure 5), our model was able to successfully detect changes in wild boar relative abundance. However, model predictions at this level seems to indicate the upper limit of HY data, since wild boar numbers were overestimated. This effect was expected, due to the assumptions in downscaling approaches commented above.



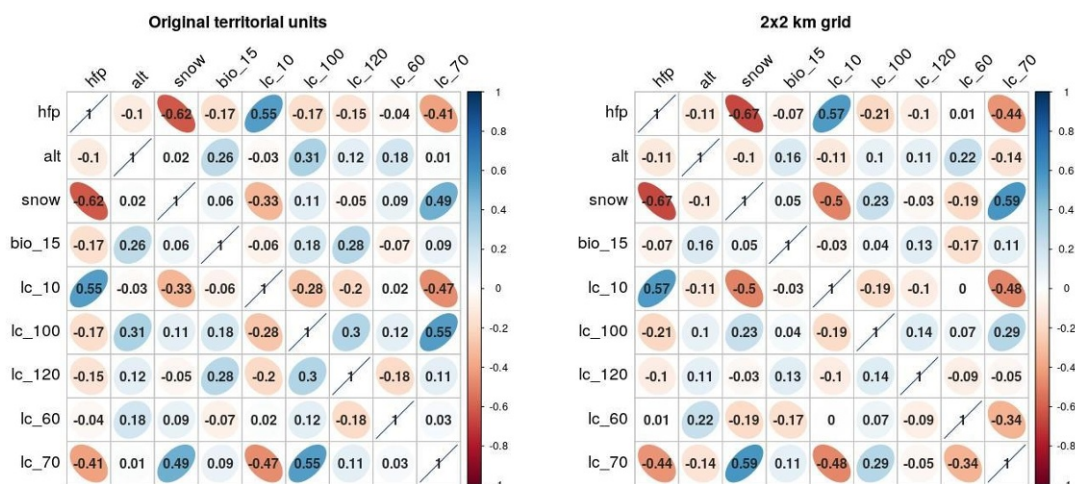
**Figure 5:** 10x10 km downscaled hunting yields predictions (cell ~ municipality) aggregated at NUTS3 level for Spain.

Overall, our results suggest a high performance and moderate accuracy of the predictions at 10x10 km grids when they are evaluated against external data (external validation). This fact can be interpreted in the light that both the spatial pattern and the predicted values at this scale can be used with confidence as a proxy of the upper limit of wild boar abundance at European scale. However, how these values relate to real wild boar density requires further research.

#### 4.2. Downscaling at 2x2 km and evaluation

Before transference, some methodological considerations should be taken into account (see Werkowska et al. 2016). First, the correlation matrix between predictors in the 2x2 km dataset should be related to that observed in the training dataset (e.g. Jiménez-Valverde et al. 2011). In this respect, correlation matrices between predictors presented, in the most cases, the same pattern for both datasets, with a strong correlation among correlation matrices (Pearson's correlation  $R = 0.98$ ; Figure 6). Only for correlations between the variables  $lc_{60} \sim lc_{70}$  and  $c_{60} \sim snow$  an inverse relation between variables was observed in the transference comparing to the training dataset, although those changes were not relevant (Pearson's  $R$  change  $< 0.4$ ; Figure 6). Only for land cover variables could some variation be expected, since these variables are categorical in the original source, and the presence of certain land cover variables excludes the presence of the others given the nature of data sources. Since the downscaling process decrease the size of cell area, there are more possibilities that cells were occupied by a single cover type, and therefore negative correlations among these categorical predictors could increase. However, due to the high number of land cover classes contemplated in the used dataset (ESA/CCI-LC

project, n =37), these correlations are probably attenuated, and non-significant ones were observed.

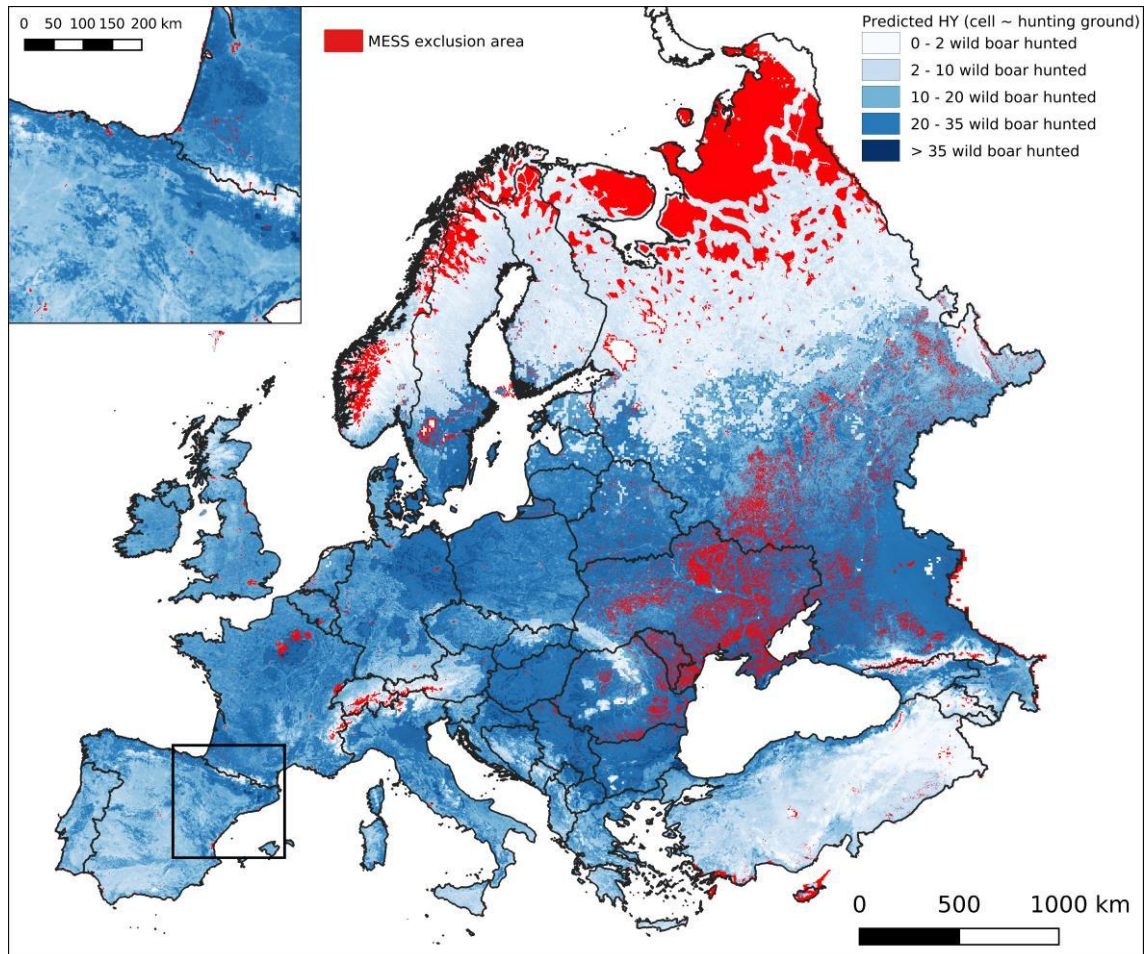


**Figure 6:** Correlation matrix between predictor variables at original territorial units used to model calibration and at 2x2 km grid used in the present report for downscaling. Note that in the most cases correlation relationships are maintained between databases.

According to MESS analyses, most of the environmental predictors in the study area at 2x2 km grid were included within the range used for model calibration. Only some areas in North Europe and other scattered locations in East Europe (Ukraine, Romania, Moldova, etc.) were not suitable to make predictions mostly due to the high altitude and snow cover values (Figure 7).

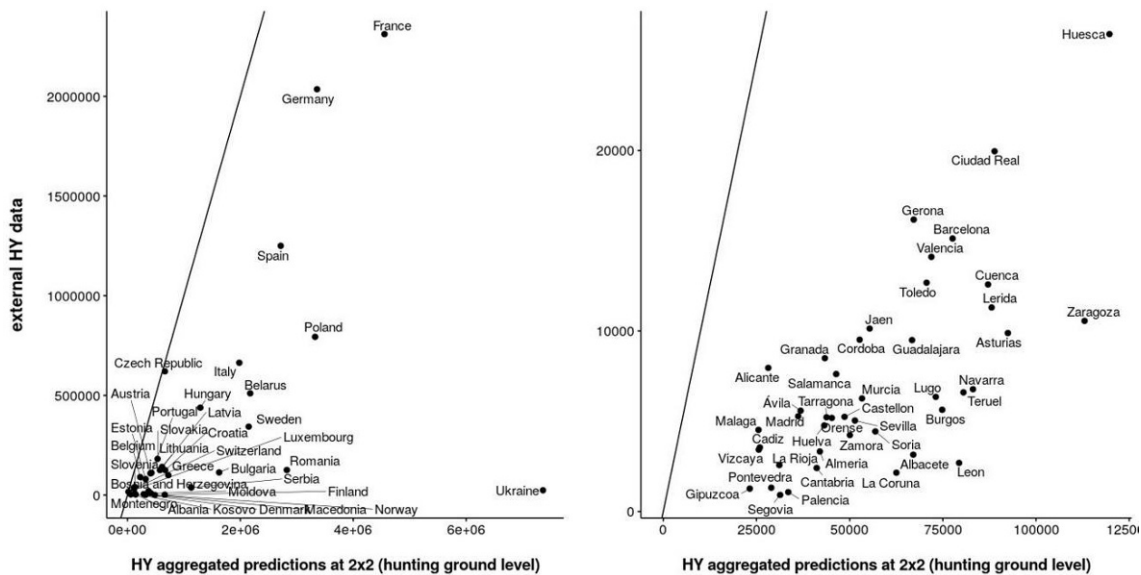
After the validation of the methodological assumptions, the model was projected on 2x2 km grid (Figure 7). In general, the pattern on the 10x10 km grid obtained in the previous report was maintained for the 2x2 km resolution. Central and East Europe showed the highest wild boar hunting yields, while South and west Europe presented moderated hunting bags. Scandinavian countries (Norway, central and North of Sweden and Finland) and north of Russia reported the lowest wild boar abundance, as well as the most of Turkey. In the downscaling procedure at 2x2 km grid, differences in HY predictions between adjacent bioregions have been slightly increased, especially in the boundaries between North and West and East and West bioregions (Figure 7). These abrupt changes when downscaling could be related to the model interaction between AREA and BIOREG. Higher prediction performance has been reported when independent models are fixed for each Bioregion (Acevedo et al. 2014; Pittiglio et al. 2018). Therefore, new modelling approaches along this line should be explored in future reports when more data at finer spatial resolution can be compiled, mainly for eastern and northern bioregions.





**Figure 7:** Map of wild boar hunting yield predicted at 2x2 km grid from the model reported in ENETWILD consortium et al. (2019). Red squares represent squares beyond the environmental domain of the model according to Multivariate Environmental Similarity Surface analysis (see Elith et al. 2010).

Similarly to 10x10 km downscaling, predictions at 2x2 km cells, and hunting yield external data, showed high model performance at countries ( $R=0.55$ ;  $p < 0.01$ ) and Spanish NUTS3 ( $R=0.7$ ;  $p < 0.01$ ) administrative levels (Figure 8). In the same way, downscaling at 2x2 km predictions overpredicts external HY data, delimiting the upper limit of the maximum wild boar abundance. In this case, overpredictions are detected at both, countries and NUTS3 levels, although the spatial pattern and linearity in the predictions is maintained. In any case, the relation between downscaled predictions and absolute HY values need to be analyzed more thoroughly in further investigations.



**Figure 8:** Relationship between aggregated model predictions at 2x2 km grid cells at different administrative levels and external wild boar hunting data Left: EU; Right: Spain. Lines show the identity ( $y=x$ ).

#### 4.3. Conclusions and further steps

The conclusions from the current study are:

- Hunting yield model downscaled at 10x10 km grid (assuming each cell corresponds to municipality level) has a good predictive performance when it was assessed on external data at lower resolution such as NUTS3 and countries;
- The correlation pattern among predictors retained in the hunting yield model does not mostly change when a finer 2x2 km cells are considered, therefore, the model is able to be downscaled at finer spatial resolutions. Hunting yield model is able to produce a relative abundance index of wild boar abundance at 2x2 km spatial resolution and European scale;
- According to MESS analyses, most of the environmental range of predictors at 2x2 km cells is within the environmental domain of the hunting yield model;
- Model predictions at both 10x10 km and 2x2 km grids overestimated the obtained hunting bage data (set). The reason for this could be that in these model predictions it is assumed and considered that the whole territory is available for hunting. However, we have not yet enough information to estimate a ~~correction~~ factor enabling to correct for this overestimation, since precise information of population abundance is not yet available at large spatial scales and should be evaluated in further research.
- At both scales, model predictions maintain a significant relationship with external hunting yield data used for evaluation, therefore they can be considered as a reliable abundance index for wild boar.

Next steps of the modelling exercise:

- In this report we have identified different areas for which the hunting yield model may have overpredicted wild boar abundance. These regions could be related to wild boar expansion areas, or to low quality datasets (unreported data), especially in the Balkans

area. Therefore, updated wild boar hunting yield data, at the finest spatial resolution as possible, is needed to correctly recalibrate our model at these regions.

- The need of finer spatial resolution hunting yield data for eastern countries is also raised by MESS analyses. Using raw data at municipality or hunting ground level likely enlarge the environmental domain of the model and therefore its capability to predict at 2x2 km cells in these areas.
- Alternative modeling approaches should be addressed in order to smooth differences in wild boar abundance between bioregions (e.g. independent models for each bioregion, etc.).
- Some efforts should be focused in relating hunting yield data/predictions with absolute values of wild boar density in order to be correctly interpreted in the risk analyses. The inclusion of new presence probability models as a predictor variable in hunting yield models could help to control overestimations.

## References

- Acevedo, P., F. Quirós-Fernández, J. Casal, and J. Vicente. 2014. Spatial distribution of wild boar population abundance: Basic information for spatial epidemiology and wildlife management. *Ecological Indicators* 36:594–600.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control* 19:716–723.
- Alexander, N. S., G. Massei, and W. Wint. 2016. The European Distribution of *Sus Scrofa*. Model Outputs from the Project Described within the Poster – Where are All the Boars? An Attempt to Gain a Continental Perspective. *Open Health Data* 4:e1.
- Barasona, J. A., Latham, M. C., Acevedo, P., Armenteros, J. A., Latham, A. D. M., Gortazar, C., et al. 2014. Spatiotemporal interactions between wild boar and cattle: implications for cross-species disease transmission. *Veterinary research* 45(1): 122.
- Bivand, R., T. Keitt and B Rowlingson. 2019. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-3. <https://CRAN.R-project.org/package=rgdal>
- Box, G. E., and Meyer, R. D. 1986. An analysis for unreplicated fractional factorials. *Technometrics* 28(1): 11-18.
- Cameron, A. C., and P. K. Trivedi. 2013. Regression analysis of count data. Volume 53. Cambridge university press.
- Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species. *Methods in ecology and evolution* 1:330–342.
- ENETWILD-consortium, P. Acevedo, S. Croft, G.C. Smith, J.A. Blanco-Aguilar, J. Fernandez-Lopez, et al. 2019. ENETwild modelling of wild boar distribution and abundance: update of occurrence and hunting data-based models. EFSA Supporting Publications, 16:1674E.
- Hijmans, R. J. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93:679–688.
- Jiménez-Valverde, A., Decae, A. E., & Arnedo, M. A. 2011. Environmental suitability of new reported localities of the funnelweb spider *Macrothele calpeiana*: an assessment using potential distribution modelling with presence-only techniques. *Journal of Biogeography* 38(6): 1213-1223.
- Jolliffe, I. T. 1972. Discarding variables in a principal component analysis. I: Artificial data. *Applied statistics* 160–173.
- Keil, P., Belmaker, J., Wilson, A. M., Unitt, P., & Jetz, W. 2013. Downscaling of species distribution models: a hierarchical approach. *Methods in Ecology and Evolution* 4(1): 82-94.
- QGIS Development Team 2019. QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>
- Pearce, J., and S. Ferrier. 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation* 98:33–43.
- Pepin, K. M., Davis, A. J., Beasley, J., Boughton, R., Campbell, T., Cooper, S. M., et al. 2016. Contact heterogeneities in feral swine: implications for disease management and future research. *Ecosphere* 7(3): e01230.
- Pittiglio, C., S. Khomenko, and D. Beltran-Alcrudo. 2018. Wild boar mapping using population-density statistics: From polygons to high resolution raster maps. *PloS one* 13:e0193295.
- Royle, J. A. 2006. Site Occupancy Models with Heterogeneous Detection Probabilities. *Biometrics* 62:97–102.
- Werkowska, W., A. L. Márquez, R. Real, and P. Acevedo. 2016. A practical overview of transferability in species distribution modeling. *Environmental reviews* 25:127–133.
- Woodroffe, R. 1999. Managing disease threats to wild mammals. In *Animal Conservation Forum* Vol. 2, No. 3, pp. 185-193. Cambridge University Press.
- Zuur, A. F., E. N. Ieno, and C. S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1:3–14.

## Glossary

- **Absolute population density:** see population density
- **Abundance estimate :** The number of individuals in a population calculated by statistical methods.
- **AICc:** A second-order Akaike's information criterion, necessary for small samples.
- **ASF:** African Swine Fever.
- **AUC:** Area Under Curve. Refers to the area under a Receiver Operator Curve (ROC) plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The resulting value reflects the predictive accuracy of a model where 0.5 indicates predictions no better than random (i.e. uninformative) and 1 indicates perfect prediction. Typically, values of 0.8 or greater are considered an indication of good performance.
- **Background data:** When not reliable information about species absence is available, a set of random points (background data) within the model extent is used to model species presence in relation to the overall environmental conditions present in the area, that are reflected in the background data.
- **Bayesian:** Method of statistical inference based on Bayes' Theorem used to update hypothesis as new information becomes available.
- **Bioregion:** Homogeneous bioclimatic regions based on bioclimatic variables, vegetation cover and topographic covariates associated to wild boar density
- **Catch per unit effort (CPUE):** Number of individuals harvested through hunting per unit of hunting effort, for example, number of hunter days. Described standardised hunting bag counts which reflect relative density.
- **Cross-validation:** Method of evaluating predictive models by partitioning sample data into a training set to fit the model and a testing set to use for evaluation.
- **Downscaling:** It is a procedure to obtain predictions from a statistical model at a higher spatial resolution than used to parameterize the model. In this case, models were transferred from NUTS3 to UTM 10x10 km resolution.
- **Ensemble:** Framework incorporating the output from a suite of different models.
- **Environmental domain:** The range of environmental predictors that is included in the training datasets. That is, if you train a model within a range 2-20°C of temperature, the model only is able to explain the response to the species to that range, but the model does not have information about how the species is able to respond in localities without that range.
- **Environmental potential:** It is term closely related with potential distribution. It shows where the environmental conditions are similar to those present in the species distribution range.
- **GBIF:** Global Biodiversity Information Facility.
- **GLM:** Generalised Linear Model.
- **Habitat suitability:** the ability of a habitat to provide a species life requisites under current conditions.
- **Hunting bag:** It refers to the number of animals hunted in a territory usually during a given hunting season.

- **Hunting yield:** It is usually used to refer to a relative abundance index based on hunting bag data.
- **IUCN:** International Union for Conservation of Nature.
- **MESS:** Multivariate Similarity Surface.
- **Model extent:** This term refers to the geographical area on which the model is to be fitted.
- **NUTS3:** Nomenclature of Territorial Units for Statistics Level 3.
- **Population density (d):** It is a measurement of population size per area unit, i.e., population size divided by total land area. The absolute density usually is expressed in heads per 100 ha. Multiplying the population density by the studied surface, we obtain the population size. It can be calculated by different methods (either direct or indirect, summarized in Table 2).
- **Population size or absolute abundance (N):** It is the size of the population. It can be a known or estimated number, expressed in number of individuals. When related to area unit it gives the population density.
- **Potential distribution:** The range in which there are favourable abiotic conditions for the occurrence of a species.
- **Predicted relative abundance:** Refers to how common or rare a species is in a defined location expressed as relative abundance as a result of quantifying the effect of a group of predictors (predictive modelling) on the relative abundance of such species (normally, at large spatial scale).
- **Predictive accuracy:** Quantitative metric describing the accuracy of model predictions. Computed by comparing model predictions against independent data often obtained through a process of cross-validation.
- **Presence-absence:** Dataset contain independent locations with binary classification describing whether a species is present or explicitly absent.
- **Presence-background:** Independent datasets describing environmental conditions at locations where a species has been observed and those of a random sample from the available landscape.
- **Presence-only:** Dataset containing independent events (date, location, recorder) describing species sightings; positive occurrences.
- **Probability of detection:** Probability of detecting a species at a particular location given it is present.
- **Probability of occurrence:** Probability of a species being present at a particular location.
- **Pseudo-absence:** Unconfirmed absences selected at random according to a set of protocols from locations where a species has not been observed.
- **Realised distribution:** The actual distribution range of a species.
- **Relative abundance:** This is a relative index related with the population size (i.e. number of individuals in the population) or its density, but that is not able to estimate these absolute values.
- **Relative abundance:** Index describing the difference in populations across locations. Typically expressed using a discrete classification scale. When expressed as a continuous scale relative abundance can be transformed in absolute abundance using a population count at a single location.
- **Relative score:** Index describing the difference in suitability, i.e. likelihood of species presence, across locations.

- **RSF:** Resource Selection Function.
- **SDM:** Species distribution model.
- **Spatial autocorrelation:** Measures the degree to which observations at nearby locations are similar to each other.
- **Suitability:** Measure of how suitable a location is for a particular species; analogous to the likelihood that a species is present.
- **Training dataset:** Split the dataset is a common modelling practice aimed to use a proportion of data to fit the model (training dataset) and the rest of data to assess the model performance on independent (i.e. not use in model fitting) data (evaluation dataset).
- **TSS:** True skill statistic. Maximised sum of sensitivity (proportion of correctly predicted presences) and specificity (proportion of correctly predicted absences).
- **Variable importance:** Quantitative measure of the relative importance/contribution of model variables in explaining observed data.
- **WBDM:** ENETWILD Wild Boar Data Collection Model.